Final Technical Report

USGS Award Number: G17AP00007
Title: Digitization of Harvard-Adam Dziewoński Analog Seismograms from 1933 to 1953 with a Focus on Caribbean Earthquakes

Miaki Ishii
Department of Earth & Planetary Sciences
Harvard University
20 Oxford Street
Cambridge, MA 02138
(617)384-8066
(617)384-8249 (fax)
ishii@eps.harvard.edu

Term: January 1, 2017 to December 31, 2017

Abstract

The goals of this project can be divided into two. The first is to improve the digitization software, DigitSeis, so that it is more robust and accessible, and the second is to use the software to digitize some HRV recordings of the Caribbean earthquakes that occurred between 1933 and 1953. We have made significant progress in the capabilities and usability of the DigitSeis software. It is practically a different software now compared to 2016. We have also digitized 14 images of the HRV long-period recordings between 1935 and 1952 that contain seismic wave arrivals from events in the Caribbean. The extracted SAC files are made available online at our web pages along with the new versions of the DigitSeis software.

Report

The main aims of the project was to attempt to digitize seismograms from the Harvard-Adam Dziewoński Observatory with recordings of earthquakes from the Caribbean region between 1933 and 1953, and to continue our development of a software that allows robust and highly accurate conversion of the analog images to digital time series. With this grant, we have made significant progress on both aims as described below.

DigitSeis Software

PI Ishii took over the development of this software from a former postdoc early January of 2017, and has been implementing changes. Just to illustrate the magnitude of the work that went in, the software that was released (v0.53) prior to the PI involvement consisted of about 4,000 lines of code. More than 80% of this program has been either modified, replaced, or removed, and the current version (v1.3) contains more than 24,000 lines of code. With the implementations of major changes that are highlighted below, we have released newer versions online at
http://www.seismology.harvard.edu/research/DigitSeis.html
once in June of 2017 (v1.1), and again in March of 2018 (v1.3). We are in the process of writing a short report on version 1.3 of the software and its capabilities (with focus on testing its accuracy and discussing limitations) to be submitted to the *Seismological Research Letters*.

- **Stand-Alone Processing:** The initial version of the software included a handful of tools to clean the image (e.g., remove stains, adjust contrast), but all of these functionalities applied to the entire image, rather than sections of the image. Analog seismograms often have regions of better/poor exposure, as well as better/poor conditions, and full-image processing is not ideal. The only solution used to be that the image must first be processed using another software (e.g., Photoshop) before it was imported into DigitSeis. This has been changed in two ways. The classification and digitization algorithms no longer require cleanest possible images to work (in fact, the new version uses the original image, including stains and hand-written notes, for calculation), hence a third-party software is not needed. Another improvement is automation of image cleanup, and when this is not sufficient, there are tools that allow the user to work on small portions of the image. This significantly reduced the workload, and contributed to speed up of digitizing images. Furthermore, some images that were untreatable with the earlier version can now be classified and digitized.

- **Efficiency:** Many of the algorithms have been rewritten to improve efficiency of the program. Two examples are in object classification and digitization (two of the most important components of the software). The classification process used to be very cumbersome, with even a simple change such as changing an object type (e.g., to make an object that has been misclassified as noise into a time mark) easily taking 20–30 seconds, and in the worst case, about a minute. From the user's point of view, this is frustrating; after clicking twice, the user needs to wait for so long before the next object can be processed. Now, the change is nearly instantaneous. Other features needed in object classification (e.g., region removal, undo) have sped up in a similar manner. Another example is the digitization process. For simple images, it used to take about 15 minutes, and for complicated ones, it took up to an hour. With a different approach and a new algorithm, it is now finished in less than a minute, even for the most complicated images that have been processed so far.

- **Automation:** With the vast amount of seismograms available to be digitized, any automation is desirable. The code has, therefore, been scrutinized for any possibility of automation, and when possible, automated. For example, timing calculation used to require the user to manually select the starting positions of each trace, ending positions of each trace, and the positions of the first and last time marks of each trace. All of these have been automated (with the option for the user to make modifications if automation resulted in imperfect location determinations) so that it is done with a single button click.

- **Robustness:** The original version of DigitSeis had quite a number of issues, both in terms of the setup and algorithms. For example, user could inadvertently open multiple analyses, or if there are multiple windows (e.g., main DigitSeis window and classification analysis window) and mistakenly clicked on the wrong window, the next figure would appear in the wrong place. These issues have been corrected in the new version. Another major problem with the previous version was that some algorithms resulted in errors that either stopped the code (the user needed to close the analysis and reopen it), slowed the processing down, or produced numerous error messages (but did not have functional problems). All these known bugs have been fixed in the new version, and the software now produces practically no error messages, and the user is not forced to unnecessarily terminate the program. The third way in which the code has been improved is to take into account unusual traces that have problems when using default parameters. For example, there are images for which the detection of trace-zero lines do not work with the default setting. It used to be that in such cases, the user needed to make changes to the source code, and try another number. These issues have been changed in two ways. One is to provide tools so that the user can modify default settings (more below), and another is to take these possibilities into account in the algorithms. The final type of modification that improved robustness of the code is the conversion into more mistake-friendly code. Even when the user makes mistakes, in most cases now, the user can

relatively easily correct it. In the old version, it often resulted in major parts of the analysis that needed to be redone.

- **Accuracy:** Accuracy of the digitization process has been improved through three approaches. One is to implement algorithms that are more accurate. For example, the trace-zero lines used to be calculated using rough estimates based upon image intensity distribution. The new version takes advantage of some of the processing that has already been done, and uses the trace objects to determine the trace-zero line. This approach required some rethinking of the processes that are done, and reorganization of the program, but produces more reliable results. The second way in which the code improves accuracy is by providing the user with check tools. For example, nearly 3000 objects exist on a single analysis, and it is difficult to make sure that every object is classified correctly (e.g., no noise object is mistakenly classified as time marks). The software now includes an ability to display locations of the time marks so that the user can easily check if any objects have been misclassified. Similarly, there is a check algorithm that allows user to see if any section of the trace has been missed. In addition to improving accuracy of the analyses, these tools resulted in a significant reduction in time for a single seismogram analysis. Finally, there are additional tools that allow the user to correct if any automatic calculations are not perfect. For example, in the old version, once the location of each minute position is determined by the time calculation algorithm, it was impossible to change them even if the code produced less than satisfactory positions. Now, the user has the ability to move the minute position bars so that they can be updated easily.

- **Compactness:** The method by which the analysis is saved has been completely rewritten to improve structure in which data are stored, and also to reduce the file size (and hence time needed to save or open files). For example, a complicated analysis with hundreds of segments requiring manual correction (i.e., the amplitudes are such that traces are touching those above and below), resulted in files that were more than 6 GB using the original algorithm and took tens of minutes to save. Now, the same analysis can be saved in a file that is about 700 MB, and the significant reduction in size leads to quicker save and load processes.

- **User Control:** Many of the features and default setups were hard-wired into the original version of DigitSeis, and the user had no control over them. For example, the default color scheme for the classified objects (i.e., white for main trace, green for time marks, and red for noise objects) in v0.53 does not work for people with color blindness. The default color scheme has been changed to take this into account, but the user has the freedom to set his/her favorite colors and save these as the default setup. Similarly, the user now has control over various default setups (e.g., size of the window for correcting traces) that can be easily accessed and saved.

- **Clearer Work Flow:** The layout of the original version of DigitSeis was such that the user had to know exactly which buttons needed to be employed next. It has been changed considerably in the new version in two ways. One is that the buttons have been organized so that they need to be pressed from top to bottom as the user makes progress. The other is activation and inactivation of these buttons. Many of the functions do not become available until the user completes the previous step, and some of the functionalities are turned off after user completes it. With this arrangement, someone learning the program requires much less time in getting started than in the previous version.

- **Simplification:** The new version of DigitSeis window is much simpler in appearance. Unnecessary plots, messages, and control buttons have been removed while some of the new and old functionalities that are typically not needed have been moved and hidden from the main window. Furthermore, improved programming reduced the amount of interaction with the user. For example, selecting and inputting a point position used to require the user to click on

the cursor button to activate the cursor, click on the screen, find out the x and y positions of the cursor location, manually type in the position values into a window, click on a button to get the information stored, and then manually remove the data cursor object that remained on the figure window. In the new version, a need for this interaction has been mostly removed (through automation), but if it becomes necessary, the user can click on a button to activate a point selection algorithm, click on the screen close to the desired location, move the point around until the user is satisfied, and then double-click on the symbol to finalize the position and to have it removed from the screen. These updates are not necessarily improvements to the digitization process itself, nonetheless significantly reduce the amount of time the user needs to spend on a single trace.

- **Saved Information:** Even though the previous version of DigitSeis generated files that were several factors larger in size, some of the important information were not saved (while some unnecessary data were). This made repeating some processes difficult, if not impossible, and required redundant actions if the analysis was saved and closed. For example, association of each object with a specific trace was not saved, so whenever the user needed to digitize again, the association needed to be recalculated. Furthermore, manually corrected trace information was also not saved, so running automatic digitization again meant the manual correction got lost. The new version allows more flexibility by saving carefully selected data that allows user to start exactly from where he/she finished, and to have all necessary information for future analysis.

- **Seismogram Timing:** The approach for timing the digitized seismograms has been completed changed, and new timing algorithm has been implemented. This requires much less input from the user (i.e., more automatic), allows easy way for the user to check the resulting timing information (there used to be none), and is more accurate.

Caribbean Earthquake Recordings

As the DigitSeis software was being developed, it was also used to digitize recordings of Caribbean earthquakes between 1935 and 1952 using the analogue seismograms from the HRV station. Due to simpler waveforms and lower noise level, the long-period seismograms are typically easier to work with than short-period recordings, and so we focused on the long period. The images that were digitized correspond to events summarized in Table 1. For most of the images, the timed SAC files were successfully extracted, and are made freely available online at the yearly pages that are linked from
http://seismology.harvard.edu/HRV/scanned_images.html
One of the components that we analyzed, however, had unusual time marks (indicated by having thicker lines), and the current version of DigitSeis was not successful in assigning appropriate time to the digitized traces (component indicated with red color in Table 1). This will be something that we will look into as the software improves in the future. Once the record can be reliably timed, it is out intention to upload the extracted SAC files so that anyone interested in studying the events have free access to them.

Table 1: Summary Digitized HRV Recordings Containing Caribbean Earthquakes

Date and Time	Latitude	Longitude	Depth	Magnitude	Digitized Components
19351110 18:27:00	16.500	-62.500	100.0	6.2	EW Z
19390612 04:05:00	20.500	-66.000	35.0	6.2	NS EW Z
19400706 03:40:00	13.000	-61.250	160.0	6.5	EW
19460325 08:47:00	19.750	-74.750	35.0	6.0	Z
19460521 09:16:46	14.822	-60.571	30.0	6.9	NS EW Z
19461004 14:45:00	18.750	-68.500	50.0	7.0	NS EW Z
19470807 00:40:28	19.600	-75.312	35.0	6.7	Z
19521028 04:29:59	18.504	-73.797	25.0	6.0	NS

A table summarizing the ISC hypocentral information of earthquakes with magnitude ≥ 6.0 that were digitized as part of this grant. The Date and Time column gives the date (YYYYMMDD format) and time (HH:MM:SS, UTC time). Whenever the ISC catalogue included multiple magnitude estimates from various contributors, the average value is shown. The Digitized Components column indicates the long-period components that have been processed (EW for east-west, NS for north-south, and Z for vertical). The component in red has been traced out, but due to difficulties in using time marks, has not been converted to SAC format and hence are not made available online.